



Tarea Online SAA02

Realizado por: André Manuel Santamaría Regal

- **Apartado 1: Analiza los siguientes casos prácticos de proyectos en los que se quiere aplicar modelos de Aprendizaje Automático (Machine Learning) y determina:**

- Qué tipo de Aprendizaje habría que aplicar (supervisado, no supervisado o por refuerzo).
 - Qué tipo de algoritmo o algoritmos se podría/n utilizar para obtener resultados óptimos.
-

- **Caso 1: Para un hospital.**

- Queremos obtener un modelo que ayude a detectar posibles casos de diabetes entre mujeres de cualquier edad.
- Nuestra base de datos consta de 156.374 fichas con diferentes casos ya estudiados.
- Tenemos 8 campos de datos numéricos:
 - Embarazos.
 - Glucosa plasmática.
 - Presión de la sangre.
 - Grosor de la piel en el tríceps.
 - Nivel de insulina.
 - BMI ().
 - Pedigrí de diabetes.
 - Edad.
- Y también conocemos para cada caso si tuvieron Diabetes o no (dato tipo categoría Sí/No).

En este 1er caso aplicaría un tipo de aprendizaje supervisado, ya que disponemos de una base de datos con historiales médicos ya estudiados. El algoritmo que aplicaría sería el de un árbol de decisiones, aunque al ser el resultado un dato de si o no se podría probar también con un algoritmo de regresión logística, al ser las dos un tipo de clasificación binaria ya que buscamos predecir algo que tiene dos posibles valores (diabetes si o no).

- **Caso 2: Para una empresa del sector inmobiliario.**
 - Queremos obtener un modelo que prediga el precio de venta de un inmueble.
 - Nuestra base de datos consta de 239.751 inmuebles vendidos en el último año en todo el territorio nacional.
 - Contamos con la siguiente información por cada inmueble:
 - Metros cuadrados habitables.
 - Número de habitaciones.
 - Categoría de inmueble (piso, casa, adosado).
 - Código postal donde se encuentra el inmueble.
 - Valor según catastro.
 - Año de construcción.
 - Número de reformas realizadas hasta la fecha.
 - Precio de venta.

En este 2º caso también usaría aprendizaje supervisado, ya que disponemos de una base de datos con información veraz sobre la venta de inmuebles en España el último año. En cuanto al algoritmo utilizaría la regresión lineal ya que necesitamos predecir o pronosticar el valor de una vivienda teniendo en cuenta unos datos cuantitativos que ya poseemos.

- **Caso 3: Para una entidad bancaria.**
 - Queremos obtener un modelo que ayude a detectar posibles usos fraudulentos en tarjetas de crédito.
 - Nuestra base de datos cuenta con 2.351.298 datos de transacciones en los últimos nueve meses.
 - Para cada transacción contamos con estos datos:
 - Fecha.
 - Hora.
 - Localización (código postal) donde se ha realizado la transacción.
 - Cantidad de dinero que se ha gastado en la transacción.
 - Tipo de dispositivo en el que se ha realizado la transacción (cajero, comercio de productos, comercio de servicios).
 - Saldo medio del propietario de la tarjeta en el último año.
 - Saldo actual del propietario de la tarjeta en el momento de hacer la transacción.

En este 3er caso tenemos un caso de aprendizaje no supervisado, ya que aunque se dispone de una base de datos con datos etiquetados no disponemos la etiqueta de si es fraude o no, como algoritmo utilizaría el algoritmo K-means ya que una de sus utilidades es la detección de datos anómalos y un uso fraudulento de una tarjeta se puede detectar por ser un uso anómalo.

- **Caso 4: Para un fabricante de teléfonos móviles.**
 - Queremos incorporar en la cámara de fotos del móvil una tecnología que reconozca caras para que en esos casos la cámara las enfoque prioritariamente al resto de elementos que aparezcan en el encuadre.
 - Contamos con un banco de más de 2.000.000 de imágenes clasificadas y etiquetadas en las que se indica en cuáles hay rostros humanos, y, en las que sí los hay, en qué zonas de la imagen se encuentran.

Este 4º caso se trata de un problema de aprendizaje supervisado ya que disponemos de datos etiquetados con la variable objetivo (localizar rostros humanos en imágenes). Emplearía un algoritmo de tipo SVM ya que es muy adecuado para resolver problemas de clasificación, en este caso de imágenes.